09/402488
PCT/CA98/00398
15 JUNE 1998 (15.06.98)

REC'D 14 JUL 1998
WIPO PCT

THE UNITED STATES OF AMERICA

TO ALL TO WHOM THESE PRESENTS SHALL COME:

UNITED STATES DEPARTMENT OF COMMERCE
United States Patent and Trademark Office

May 28, 1998

THIS IS TO CERTIFY THAT ANNEXED HERETO IS A TRUE COPY FROM THE RECORDS OF THE UNITED STATES PATENT AND TRADEMARK OFFICE OF THOSE PAPERS OF THE BELOW IDENTIFIED PATENT APPLICATION THAT MET THE REQUIREMENTS TO BE GRANTED A FILING DATE UNDER 35 USC 111.

APPLICATION NUMBER: ***60/044,254***
FILING DATE: ***April 25, 1997***

PRIORITY DOCUMENT

SUBMITTED OR TRANSMITTED IN
COMPLIANCE WITH RULE 17.1(a) OR (b)

By Authority of the
COMMISSIONER OF PATENTS AND TRADEMARKS

D. S. Butler

D.S. BUTLER
Certifying Officer

2

A / Prov

65114 U.S. PTO
04/25/97

11105 U.S. PTO
60/044254
04/25/97

Our Ref.: 9369-40

IN THE UNITED STATES PATENT AND TRADEMARK OFFICE

Assistant Commissioner for Patents
BOX PROVISIONAL
Washington, D.C.
U.S.A. 20231

Dear Sir:

PROVISIONAL APPLICATION UNDER 37 C.F.R. § 1.53(b)(2)

This is a request for filing a Provisional application under 37 C.F.R. § 1.53(b)(2) entitled **Method for Cleavage of Fusion Proteins** by the following inventor(s):

Full Name of Inventor 1st	Family Name Moloney	First Given Name Maurice	Second Given Name
Residence & Citizenship	City Calgary	State or Foreign Country Alberta CAX	Country of Citizenship Irish
Post Office Address	Post Office Address 34 Edgebrook Cove, N.W.	City Calgary	State & Zip Code/Country Alberta, Canada T3A 5N5

Full Name of Inventor 2nd	Family Name Alcantara	First Given Name Joenel	Second Given Name
Residence & Citizenship	City Calgary	State or Foreign Country Alberta CAX	Country of Citizenship Canadian
Post Office Address	Post Office Address 3 Castledale Place N.E.	City Calgary	State & Zip Code/Country Alberta, Canada T3J 1Y4

Full Name of Inventor 3rd	Family Name van Rooijen	First Given Name Gijs	Second Given Name
Residence & Citizenship	City Calgary	State or Foreign Country Alberta CAX	Country of Citizenship Dutch
Post Office Address	Post Office Address 3223 Bearspaw Drive N.W.	City Calgary	State & Zip Code/Country Alberta, Canada T2L 1T1

1. __X__ Enclosed is the Provisional application as follows: __14__ pages of specification, __8__ claims, __1__ page Abstract and __2__ sheets of drawings.

2. ____ Enclosed is a Verified Statement that this filing is by a small entity (37 C.F.R. 1.9, 1.27, 1.28).

3. __X__ Payment of Provisional filing fee under 37 C.F.R. § 1.16(k):

   ____ Attached is a cheque in the amount of $_____.
   __X__ Please charge the filing fee of $150.00 from our Deposit Account No. 02-2095. The letter is being filed in triplicate.
   ____ PAYMENT OF THE FILING FEE IS BEING DEFERRED.

4. __X__ The Commissioner is hereby authorized to charge any additional fees as set forth in 37 C.F.R. §§ 1.16 to 1.18 which may be required by this paper or credit any overpayment to Account No. 02-2095.

5. ____ Enclosed is an Assignment of the invention to ________________. Recordation Form Cover Sheet and a cheque for $40 to cover the Recordation Fee.

6. ____ Also enclosed:

7. ____ The invention was made by the following agency of the United States Government or under a contract with the following agency of the United States Government: ________________________.

8. __X__ Address all future communications to the Attention of Micheline Gravelle at the address below.

Respectfully submitted,

April 24, 1997
Date

H. Samuel Frost
Registration No. 31,696

BERESKIN & PARR
Box 401
40 King Street West
Toronto, Ontario
Canada M5H 3Y2
Telephone: 416-364-7311

B&P File No. 9369-040/MG

BERESKIN & PARR **UNITED STATES**

60044254 - 042504

Title: METHOD FOR CLEAVAGE OF FUSION PROTEINS
Inventors: Gijs van Rooijen, Joenel Alcantara and Maurice M. Moloney

B&P File No. 9369-040/MG

Title: METHOD FOR CLEAVAGE OF FUSION PROTEINS

FIELD OF THE INVENTION

The present invention relates to an improved method for recovering recombinantly produced polypeptides. The method involves expressing the recombinant polypeptide as a fusion protein with a pro-peptide. The pro-peptide-polypeptide fusion protein can be cleaved and the recombinant polypeptide released under the appropriate conditions.

BACKGROUND OF THE INVENTION

Several enzymes, such as pepsin and chymosin, are synthesized *in vivo* as inactive precursors or zymogens. Zymogens become activated under appropriate conditions, which in the case of pepsin and chymosin are provided by the acidic environment of the mammal stomach. In the process, two interrelated events occur: (1) the zymogen is activated by a conformational change in the protein (activation step) and sequentially or simultaneously (2), an amino-terminal peptide, referred to in the art as the pro-peptide, is cleaved from the zymogen to form the mature protein (processing step). The activation and processing of zymogens has been extensively characterized (McCaman and Cummings, 1986, J. Biol. Chem. 261: 15345-15348) although so far the process has only been studied using the pro-peptide linked to its homologous mature protein.

An important factor for the commercial development of biotechnology is the purification of recombinant DNA products from expression systems, which typically accounts for 50% or more of the costs (Labrou, N. and Clonis, Y.D., (1994), J. Biotechn., 36 95-119). Many recombinant DNA products are produced as fusion proteins. Consequently, one step in the purification process involves an enzymatic or chemical cleavage step in which the produced protein is processed to its mature form. Very few proteolytic enzymes cleave fusion proteins specifically and efficiently. In addition, the currently

widely used proteolytic enzymes, such as blood clotting factor Xa and thrombin, are expensive, and contamination of final product with blood pathogens is a consideration.

Consequently, there is a need for an improved process for recovering recombinantly produced polypeptides from their expression systems.

SUMMARY OF THE INVENTION

The present inventors have developed a novel method for recovering recombinantly produced polypeptides. The method involves expressing the polypeptide as a fusion protein with a pro-peptide so that the recombinant polypeptide can be cleaved from the pro-peptide under the appropriate conditions.

In one aspect, the invention provides a chimeric DNA sequence encoding a fusion protein, the chimeric DNA sequence comprising a first DNA sequence encoding a pro-peptide and a second DNA sequence encoding a polypeptide that is heterologous to the pro-peptide.

In another aspect, the present invention, provides a method for the preparation of a recombinant polypeptide comprising

(a) introducing into a host cell an expression vector comprising:

(1) a first sequence capable of regulating transcription in a host cell, operatively linked to

(2) a second chimeric DNA sequence encoding a fusion protein, said chimeric DNA sequence comprising a first DNA sequence encoding a pro-peptide, linked in reading frame to a second DNA sequence heterologous to said pro-peptide and encoding said recombinant polypeptide; operatively linked to

(3) a third DNA sequence encoding a termination region functional in the host cell,

(b) growing said host cell to produce said fusion protein; and

(c) altering the environment of said fusion protein so that said pro-peptide is cleaved from said fusion protein to release said recombinant polypeptide.

Other objects, features and advantages of the present invention will become apparent from the following detailed description. It should be understood, however, that the detailed description and the specific examples while indicating preferred embodiments of the invention are given by way of illustration only, since various changes and modifications within the spirit and scope of the invention will become apparent to those skilled in the art of this detailed description.

BRIEF DESCRIPTION OF THE DRAWINGS

The invention will now be described in relation to the drawings in which:

Figure 1 is the nucleic acid and deduced amino acid sequence of a GST-Chymosin pro-peptide-Hirudin sequence.

DETAILED DESCRIPTION OF THE INVENTION

As hereinbefore mentioned, the present invention relates to a novel method for preparing and recovering recombinant polypeptides.

Accordingly, the present invention provides a method for the preparation of a recombinant polypeptide comprising:

(a) introducing into a host cell an expression vector comprising:

(1) a first sequence capable of regulating transcription in a host cell, operatively linked to

(2) a second chimeric DNA sequence encoding a fusion protein, said chimeric DNA sequence comprising a first DNA sequence encoding a pro-peptide, linked in reading frame to a second DNA sequence heterologous to said pro-peptide and encoding said recombinant polypeptide, operatively linked to

(3) a third DNA sequence encoding a termination region functional in said host cell,

b) growing said host cell to produce said fusion protein; and

c) altering the environment of said fusion protein so that said pro-peptide is cleaved from said fusion protein to release said recombinant polypeptide.

The term "pro-peptide" as used herein means the amino terminal portion of a zymogen up to the maturation site.

The polypeptide can be any polypeptide that is heterologous to the pro-peptide meaning that it can not be the mature protein that is normally associated with the pro-peptide as a zymogen.

The invention also includes a chimeric DNA sequence encoding a fusion protein, the chimeric DNA sequence comprising a first DNA sequence encoding a pro-peptide and a second DNA sequence encoding a polypeptide that is heterologous to the pro-peptide.

The chimeric DNA sequences which encode the fusion proteins of the present invention can be incorporated in a known manner into a recombinant expression vector which ensures good expression in a host cell.

Accordingly, the present invention also includes a recombinant expression vector comprising a chimeric DNA molecule of the present invention operatively linked to a regulatory sequence and termination region suitable for expression in a host cell.

The term "suitable for expression in a host cell" means that the recombinant expression vectors contain the chimeric DNA sequence of the invention a regulatory sequence and a termination region, selected on the basis of the host cells to be used for expression, which is operatively linked to the chimeric DNA sequence. Operatively linked is intended to mean that the chimeric DNA sequence is linked to a regulatory sequence and a termination region in a manner which allows expression of the chimeric sequence.

Regulatory sequences and termination regions are art-recognized and are selected to direct expression of the desired protein in an appropriate host cell. Accordingly, the term regulatory sequence includes promoters, enhancers and other expression control elements. Such regulatory sequences are known to those skilled in the art or one described in Goeddel, Gene Expression Technology: Methods in Enzymology 185, Academic Press, San Diego, CA (1990) can be used. It should be understood that the design of the expression vector may depend on such factors as the choice of the host cell to be transformed and/or the type of protein desired to be expressed. Such expression vectors can be used to transform cells to thereby produce fusion proteins or peptides encoded by nucleic acids as described herein.

The recombinant expression vectors of the invention can be designed for expression of encoded fusion proteins in prokaryotic or eukaryotic cells. For example, fusion proteins can be expressed in bacterial cells such as *E. coli*, insect cells (using baculovirus), yeast cells, plant cells or mammalian cells. Other suitable host cells can be found in Goeddel, Gene Expression Technology: Methods in Enzymology 185, Academic Press, San Diego, CA (1990).

Expression in prokaryotes is most often carried out in *E. coli* with vectors containing constitutive or inducible promotors directing the expression of the fusion proteins. Inducible expression vectors include pTrc (Amann et al., (1988) Gene 69:301-315) and pET 11d (Studier et al., Gene Expression Technology: Methods in Enzymology 185, Academic Press, San Diego, California (1990) 60-89). While target gene expression relies on host RNA polymerase transcription from the hybrid trp-lac fusion promoter in pTrc, expression of target genes inserted into pET 11d relies on transcription from the T7 gn10-lac 0 fusion promoter mediated by coexpressed viral RNA polymerase (T7 gn1). This viral polymerase is supplied by host strains BL21(DE3) or HMS174(DE3) from a resident λ prophage harboring a T7 gn1 under

the transcriptional control of the lacUV 5 promoter. Another attractive bacterial expression system is the pGEX expression system (Pharmacia) in which genes are expressed as fusion products of glutathione-S-transferase (GST), allowing easy purification of the expressed gene from a GST affinity column.

One strategy to maximize recombinant protein expression in *E. coli* is to express the protein in host bacteria with an impaired capacity to proteolytically cleave the recombinantly expressed proteins (Gottesman, S., Gene Expression Technology: Methods in Enzymology 185, Academic Press, San Diego, California (1990) 119-128). Another strategy is to alter the nucleic acid sequence of the chimeric DNA to be inserted into an expression vector so that the individual codons for each amino acid would be those preferentially utilized in highly expressed *E. coli* proteins (Wada et al., (1992) Nuc. Acids Res. 20: 2111-2118). Such alteration of nucleic acid sequences of the invention could be carried out by standard DNA synthesis techniques.

Examples of vectors for expression in yeast S. cerivisae include pYepSec1 (Baldari. et al., (1987) Embo J. 6:229-234), pMFa (Kurjan and Herskowitz, (1982) Cell 30:933-943), pJRY88 (Schultz et al., (1987) Gene 54:113-123), and pYES2 (Invitrogen Corporation, San Diego, CA).

Baculovirus vectors available for expression of proteins in cultured insect cells (SF 9 cells) include the pAc series (Smith et al., (1983) Mol. Cell Biol. 3:2156-2165) and the pVL series (Lucklow, V.A., and Summers, M.D., (1989) Virology).

The polypeptide of the present invention may be any polypeptide that is not normally fused to the pro-peptide used in the method. The polypeptide is preferentially stable under cleavage conditions, for example at acidic pH, and the polypeptide may be activated after cleavage upon adjusting the pH, or altering the environment otherwise so that conditions optimal for enzymatic activity are generated.

The pro-peptide used in the present invention may be any pro-peptide including pro-peptides derived from aspartic proteases, serine proteases, lipases, and esterases. In preferred embodiments of the invention, the pro-peptide is derived from chymosin, trypsinogen, pepsin or insulin. The invention also includes the full-length pro-peptide as well as portions of the pro-peptide or mutated forms of the pro-peptide. Mutated forms of the pro-peptide may be used to obtain specific cleavage between the pro-peptide and a heterologous protein. Mutations in the pro-peptide could alter the optimal conditions, such as temperature, pH and salt concentration, under which cleavage of a heterologous peptide is achieved (McCaman, M.T. and Cummings, D.B., (1986), J. Biol. Chem. 261:15345-15348). Depending on the pro-peptide, cleavage of the heterologous protein from various pro-peptides, will be optimal under varying different conditions. Thus the invention will be amenable to heterologous proteins which are preferentially cleaved under a variety of desirable conditions.

The DNA sequence encoding the heterologous polypeptide is preferably fused downstream of the DNA sequence encoding the pro-peptide and concatamers containing repetitive units of pro-peptide-heterologous protein may be employed.

In one embodiment, the pro-peptide is a pro-peptide derived from chymosin and the heterologous polypeptide is hirudin (Dodt et al., 1984, FEBS Letters 65:180-183). In particular, the present inventors have constructed a chimeric DNA sequence in which the DNA encoding the chymosin pro-peptide was fused upstream of the DNA sequence encoding the leech anticoagulant protein hirudin. The gene fusion (Pro-Hirudin) was expressed in *E. coli* cells. It was found that upon lowering of the pH to pH 2, and more preferably to pH 4.5, and in the presence of a small quantity of mature chymosin, the heterologously fused protein, hirudin, was efficiently cleaved from the chymosin pro-peptide.

Although the inventors have employed mature chymosin to assist in the cleavage reaction, it is conceivable that the use of other pro-peptides may not require the addition of the mature peptide in order to accomplish efficient cleavage.

Activation of the fusion protein may be *in vitro* or *in vivo*. In one embodiment, the pro-peptide is used to facilitate cleavage from proteins recombinantly produced on oil bodies as disclosed in PCT application Publication No. WO 96/21029. In this embodiment, the pro-peptide would be fused downstream of an oil body protein and upstream of the recombinant protein or peptide of interest.

In another *in vivo* application, two vectors would be introduced in the same host. In one vector expression of the zymogen or the mature protein would be controlled by an inducible promoter system. The other vector would comprise a pro-peptide fused upstream of an heterologous protein of interest. Thus it is possible to control the moment of cleavage of the peptide or protein downstream of the pro-peptide through the promoter which controls expression of the zymogen or the mature protein. Alternatively, the two expressed genes would be combined in the same vector. In preferred embodiments of this application, the pro-peptide employed is cleaved under physiological conditions.

A peptide or protein delivery application is also envisaged. In this embodiment, a therapeutic or nutritional peptide or protein would be employed as an inactive fusion protein. Activation or maturation through cleavage would only occur upon its delivery at the unique physiological conditions prevalent at the target organ or tissue. Cleavage might be enhanced by a protease specific for the peptide. This method is particularly useful for the delivery of orally ingested vaccines, cytokines, gastric lipase, peptide antibiotics, lactase and cattle feed enzymes which facilitate digestion, such as xylanase and cellulase. For example, a therapeutic or nutritional peptide or protein fused downstream of the chymosin pro-peptide might be activated in

the mammal stomach upon ingestion of chymosin or the inactive precursor form of chymosin.

The present invention is also useful in the purification process of recombinant proteins. In one embodiment, a cell extract containing an expressed pro-peptide-heterologous fusion protein is applied to a chromatographic column. Selective binding of the fusion protein to antibodies raised against the pro-peptide sequence and immobilized onto the column, results in selective retention of the fusion protein. Instead of relying on antibodies against the pro-peptide sequence, a gene encoding another immunogenic domain or a gene encoding a peptide with affinity for a commonly used column material, such as cellulose or chitin, or any other desirable tag, may be included in the gene fusion.

The following non-limiting examples are illustrative of the present invention.

EXAMPLES

EXAMPLE 1

In the first example, the protein hirudin was prepared as a fusion protein with the chymosin pro-peptide.

Construction of a pGEX-Pro-Hirudin fusion.

The fusion protein that we studied comprises the pro-peptide of calf chymosin B (Foltmann et al, 1977; Harris et al., 1982, Nucl. Acids. Res., 10: 2177-2187) fused to hirudin variant 1 (Dodt et al., 1984, FEBS Letters 65: 180-183). The hybrid gene which encoded this fusion protein was constructed using standard PCR methods (Horton et al., 1989, Gene, 77: 61-68). The DNA sequence for this Pro-Hirudin fusion was cloned into pGEX-4T-3 (Pharmacia), downstream of the gene encoding glutathion-S-transferase (GST). The complete sequence of the GST-Pro-Hirudin sequence is shown in Figure 1.

Growth of *E. coli* transf rmed with pGEX-4T-3 and pGEX-Pro-Hirudin.

Plasmids pGEX-4T-3 and pGEX-Pro-Hirudin were transformed into *E. coli.* strain DH5α to allow for high level of expression. A single colony was used to inoculate 5ml LB-amp broth. These cultures were grown overnight. One ml of each o/n culture was used to inoculate 50 ml of LB-amp broth. These cultures were grown until OD_{600}= 0.6. At this OD, IPTG (final concentration 1mM) was added to induce the expression of the GST and GST-Pro-Hirudin fusion proteins. After this induction, the cultures were grown for an additional 3 hours at 37°C. The cells were pelleted at 5000 x g for 10 minutes, and resuspended in 5 ml Tris Buffered Saline (TBS). The resuspended cells were sonicated and centrifuged at 12000 x g for 15 minutes to separate the inclusion bodies (pellet fraction) from the soluble proteins (supernatant fraction). Western blotting of both the pellet and supernation fraction indicated that under the growing conditions described above, significant amounts (5-10%) of the GST and GST-Pro-Hirudin protein were found in the supernatant fraction. The rest (90-95%) accumulated in inclusion bodies (results not shown).

Hirudin activity measurements

The supernatant fractions of both the GST and GST-Pro-Hirudin were tested for anti-thrombin activity. The samples were treated as follows: A) 20 µl supernatant + 20 µl water B) 20 µl supernatant + 20 µl of 100 mM Sodium Phosphate pH 2.0 C) 20 µl supernatant + 20 µl of 100 mM Sodium Phosphate pH 2.0 + 2 µg chymosin (Sigma) D) 20 µl supernatant + 20 µl of 100 mM Sodium Phosphate pH 4.5 E) 20 µl supernatant + 20 µl of 100 mM Sodium Phosphate pH 4.5 + 2 µg chymosin. These samples were incubated at room temperature for 1 hour. A total of 10 µl of the samples was added to 1 ml assay buffer (20 mM Tris [pH 7.5], 100 mM NaCl, 5 mM $CaCl_2$, 0.1 unit of thrombin) and incubated for 2-3 minutes before the addition of 50µl p-tos-gly-pro-arg-nitroanilide (1 mM). Thrombin

activity was measured as a function of chromozyme cleavage by monitoring the increase in absorption at 405 nm over time (Chang, 1983, FEBS Letters, 164: 307-313). The Δ_{Abs} (405nm) was determined after 2 minutes. The result of the activity measurements are indicated in Table 1.

As can be seen from Table 1, the only extract which exhibited significant anti-thrombin activity was the extract containing the GST-Pro-Hirudin fusion which was treated at pH 4.5 and supplemented with 2 µg chymosin (E). Western blotting (results not shown) indicated that apart from treatment at pH 4.5, complete cleavage was also observed when the GST-Pro-Hirudin fusion which was treated at pH 2.0 and supplemented with 2 µg chymosin. It has been well documented that unprocessed chymosin when exposed at pH 2.0, forms a pseudochymosin, before it matures into chymosin (Foltmann et al., 1977, Scand. J. Clin. Lab. Invest. 42: 65-79; Foltmann, 1992, Proc. Natl. Acad. Sci. 74: 2321-2324; McCaman and Cummings, 1988, J. Biol. Chem. 261: 15345-15348) The pseudo chymosin cleavage site is located between the Phe^{27}-Leu^{28} peptide bond and is indicated in Figure 1. The inability of the GST-Pro-Hirudin fusion, which was treated at pH 2.0 and supplemented with 2 µg chymosin, to inhibit thrombin activity can be explained by the fact that cleavage occurred at the Phe^{27}-Leu^{28} peptide bond rather than at the Phe^{43}-Val^{44} peptide bond which separates the chymosin pro-peptide from the mature hirudin. It has been well documented that (Loison et al., 1988, Bio/Technology, 6: 72-77) mature hirudin is only active when it does not have any additional amino acids attached to its native N-terminal sequence.

While the present invention has been described with reference to what are presently considered to be the preferred examples, it is to be understood that the invention is intended to cover various modifications and equivalent arrangements included within the spirit and scope of the appended claims.

All publications, patents and patent applications are herein incorporated by reference in their entirety to the same extent as if each individual publication, patent or patent application was specifically and individually indicated to be incorporated by reference in its entirety.

DETAILED FIGURE LEGENDS

Figure 1. The nucleic acid and deduced amino acid sequence of a GST-Pro-Hirudin sequence. The deduced sequence of the chymosin pro-peptide has been underlined and the deduced hirudin protein sequence has been italicized. The hirudin nucleic acid sequence was optimized for plant codon usage. The pseudochymosin cleavage site between Phe27-Leu28 and the peptide bond separating the pro-chymosin and mature hirudin (Phe 42- Val43) are indicated with an arrow (↑).

Table 1: Activity measurements f bacterial extracts c ntaining GST (Glutathion-S-transferase) and GST-Pro-Hirudin fusi ns.

Sample	Δ Abs (405nm)/2min [Test 1]	Δ Abs (405nm)/2min [Test 2]
1 unit Thrombin	0.088	0.066
A: GST	0.087	0.082
B: GST pH 2.0	0.082	0.073
C: GST pH 2.0 +2 µg chymosin	0.063	0.073
D: GST pH 4.5	0.087	0.086
E: GST pH 4.5 +2 µg chymosin	0.087	0.087
A: GST-PRO-HIR	0.076	0.071
B: GST-PRO-HIR pH 2.0	0.072	0.064
C: GST-PRO-HIR pH 2.0 +2 µg chymosin	0.066	0.070
D: GST-PRO-HIR pH 4.5	0.078	0.075
E: GST-PROHIR pH 4.5 +2µg chymosin	0.0002	0.0001
Hirudin 2µg	0.0001	0.0001

We Claim:

1. A method for the preparation of a recombinant polypeptide comprising

a) introducing into a host cell an expression vector comprising:

(1) a first sequence capable of regulating transcription in a host cell, operatively linked to

(2) a second chimeric DNA sequence encoding a fusion protein, said chimeric DNA sequence comprising a first DNA sequence encoding a pro-peptide, linked in reading frame to a second DNA sequence heterologous to said pro-peptide and encoding said recombinant polypeptide; operatively linked to

(3) a third DNA sequence encoding a termination region functional in said host cell,

b) growing said host cell to produce said fusion protein; and

c) altering the environment of said fusion protein so that said pro-peptide is cleaved from said fusion protein to release said recombinant polypeptide.

2. A chimeric DNA sequence according to claim 1 wherein the pro-peptide is derived from aspartic proteases, serine proteases, lipases, or esterases.

3. A chimeric DNA sequence according to claim 1 wherein the pro-peptide is a pro-peptide derived from chymosin, insulin, trypsinogen or pepsin.

4. A chimeric DNA sequence according to claim 1 wherein said polypeptide is hirudin.

5. An expression vector comprising a chimeric DNA sequence according to claim 1 and a regulatory sequence suitable for expression in a host cell.

6. A transformed host cell containing an expression vector according to claim 6.

7. A method according to claim 1 wherein said altering the environment comprises altering the pH to pH 2 or to pH 4.5.

8. The method according to claim 1 wherein a mature protein of the pro-peptide is added in step (c).

ABSTRACT OF THE DISCLOSURE

An improved method for recovering recombinantly produced polypeptides is described. The method involves expressing the recombinant polypeptide as a fusion protein with a pro-peptide. The pro-peptide-polypeptide fusion protein can be cleaved and the recombinant polypeptide released under the appropriate conditions.

FIGURE 1

Nucleotide sequence and the deduced amino acid sequence of an GST-Pro-Hirudin fusi n.

```
1 ATG TCC CCT ATA CTA GGT TAT TGG AAA ATT AAG GGC CTT GTG CAA CCC ACT CGA CTT CTT      60
1 M  S  P  I  L  G  Y  W  K  I  K  G  L  V  Q  P  T  R  L  L                        20

61 TTG GAA TAT CTT GAA GAA AAA TAT GAA GAG CAT TTG TAT GAG CGC GAT GAA GGT GAT AAA   120
21 L  E  Y  L  E  E  K  Y  E  E  H  L  Y  E  R  D  E  G  D  K                       40

121 TGG CGA AAC AAA AAG TTT GAA TTG GGT TTG GAG TTT CCC AAT CTT CCT TAT TAT ATT GAT  180
 41 W  R  N  K  K  F  E  L  G  L  E  F  P  N  L  P  Y  Y  I  D                      60

181 GGT GAT GTT AAA TTA ACA CAG TCT ATG GCC ATC ATA CGT TAT ATA GCT GAC AAG CAC AAC  240
 61 G  D  V  K  L  T  Q  S  M  A  I  I  R  Y  I  A  D  K  H  N                      80

241 ATG TTG GGT GGT TGT CCA AAA GAG CGT GCA GAG ATT TCA ATG CTT GAA GGA GCG GTT TTG  300
 81 M  L  G  G  C  P  K  E  R  A  E  I  S  M  L  E  G  A  V  L                     100

301 GAT ATT AGA TAC GGT GTT TCG AGA ATT GCA TAT AGT AAA GAC TTT GAA ACT CTC AAA GTT  360
101 D  I  R  Y  G  V  S  R  I  A  Y  S  K  D  F  E  T  L  K  V                     120

361 GAT TTT CTT AGC AAG CTA CCT GAA ATG CTG AAA ATG TTC GAA GAT CGT TTA TGT CAT AAA  420
121 D  F  L  S  K  L  P  E  M  L  K  M  F  E  D  R  L  C  H  K                     140

421 ACA TAT TTA AAT GGT GAT CAT GTA ACC CAT CCT GAC TTC ATG TTG TAT GAC GCT CTT GAT  480
141 T  Y  L  N  G  D  H  V  T  H  P  D  F  M  L  Y  D  A  L  D                     160

481 GTT GTT TTA TAC ATG GAC CCA ATG TGC CTG GAT GCG TTC CCA AAA TTA GTT TGT TTT AAA  540
161 V  V  L  Y  M  D  P  M  C  L  D  A  F  P  K  L  V  C  F  K                     180

541 AAA CGT ATT GAA GCT ATC CCA CAA ATT GAT AAG TAC TTG AAA TCC AGC AAG TAT ATA GCA  600
181 K  R  I  E  A  I  P  Q  I  D  K  Y  L  K  S  S  K  Y  I  A                     200

601 TGG CCT TTG CAG GGC TGG CAA GCC ACG TTT GGT GGT GGC GAC CAT CCT CCA AAA TCG GAT  660
201 W  P  L  Q  G  W  Q  A  T  F  G  G  G  D  H  P  P  K  S  D                     220

661 CTG GTT CCG CGT GGA TCC CCG AAT TCC CGG GTC GAC TCG AGC GGC CGC GCT GAG ATC ACC  720
221 L  V  P  R  G  S  P  N  S  R  V  D  S  S  G  R  A  E  I  T                     240

721 AGG ATC CCT CTG TAC AAA GGC AAG TCT CTG AGG AAG GCG CTG AAG GAG CAT GGG CTT CTG  780
241 R  I  P  L  Y  K  G  K  S  L  R  K  A  L  K  E  H  G  L  L                     260
```

FIGURE 1 (Cont'd)

```
781  GAG GAC TTC CTG CAG AAA CAG CAG TAT GGC ATC AGC AGC AAG TAC TCC GGC TTC GTC GTC   840
261  E   D   F   L   Q   K   Q   Q   Y   G   I   S   S   K   Y   S   G   F   V   V     280
                 ↑                                                           ↑

841  TAT ACC GAC TGT ACC GAG TCC GGT CAG AAC CTC TGT CTC TGT GAG GGT TCC AAC GTC TGT   900
281  Y   T   D   C   T   E   S   G   Q   N   L   C   L   C   E   G   S   N   V   C     300

901  GGT CAG GGT AAC AAG TGT ATC CTC GGT TCC GAC GGT GAG AAG AAC CAG TGT GTC ACC GGT   960
301  G   Q   G   N   K   C   I   L   G   S   D   G   E   K   N   Q   C   V   T   G     320

961  GAG GGA ACC CCA AAG CCA CAG TCC CAC AAC GAC GGT GAC TTT GAG GAG ATC CCA GAG GAG   1020
321  E   G   T   P   K   P   Q   S   H   N   D   G   D   F   E   E   I   P   E   E     340

1021 TAT CTC CAG TAA agatctaagcttgctgctgctatcgaattcctgcagcccgggggatccactagttctagagcgg 1096
341  Y   L   Q   *                                                                   344
```